

Schrama, Sebastian

**Aktuelle Entwicklungen im „Semantic Web“ -
Das Internet auf dem Weg zum Web 3.0?**

- BACHELORARBEIT -

Hochschule Mittweida – University of Applied Science (FH)

Schrama, Sebastian

**Aktuelle Entwicklungen im „Semantic Web“ -
Das Internet auf dem Weg zum Web 3.0?**

- EINGEREICHT ALS BACHELORARBEIT -

Hochschule Mittweida – University of Applied Science (FH)

Erstprüfer: Prof. Dr. Otto Altendorfer

Zweitprüfer: Dipl. Des. Hans Florian Müller-Hermelink

Berlin, 2009

BIBLIOGRAPHISCHE BESCHREIBUNG UND REFERAT

Schrama, Sebastian:

Aktuelle Entwicklungen im „Semantic Web“ -

Das Internet auf dem Weg zum Web 3.0? - 2009 - 55S.

Mittweida, Hochschule Mittweida (FH), Fachbereich Medien,
Bachelorarbeit

Referat

In meiner Bachelorarbeit widme ich mich dem Thema „Semantic Web“. Beginnend mit einer kurzen Einleitung, gehe ich dann auf die Grundlagen der Semantik ein und erkläre die dabei auftretenden Begrifflichkeiten. Mein Hauptaugenmerk liegt dann darauf, aktuelle semantische Webanwendungen oder Projekte mit dem Bezug zum semantischen Web vorzustellen und zu bewerten. Darauf folgend gebe ich einen Überblick darüber ob und wann das Semantic Web den Massenmarkt erobert und ob es überhaupt kommerziell tragfähig ist. Des Weiteren beschäftige ich mit Problemen bei der Entwicklung und der Kritik des Semantic Web und werde dabei die Machbarkeit im Allgemeinen erörtern. Zum Schluss ziehe ich dann ein Resümee und zeige die möglichen Erfolgsaussichten einer zeitnahen Umsetzung eines „intelligenteren Internets“ auf. Wird es beispielsweise irgendwann möglich sein bei einer Suchmaschine mit einer konkreten Fragestellung nach dem nächsten Blumenladen in der Nähe zu recherchieren, der rote Rosen auf Lager hat? Und wenn ja, wie lange müssen wir darauf noch warten?

INHALTSVERZEICHNIS

I	Bibliographische Beschreibung und Referat	3
II	Inhaltsverzeichnis	4
III	Abbildungsverzeichnis	6
IV	Abkürzungsverzeichnis	7
1.	Einleitung	8
2.	Die Idee des Semantic Web	10
2.1	Probleme des Internets	10
2.2	Die Lösung – Das Semantic Web	11
2.3	Die Begriffe Semantik und Syntax	13
2.4	Annotationen	14
2.5	Ontologien	15
2.6	RDF/XML	17
2.7	RDF-Schema	19
2.8	Web-Ontologiesprache (OWL)	20
3.	Beispiele von semantischen Web-Anwendungen oder Projekten mit Bezug zum Semantic Web	21
3.1	Evri	21
3.2	Hakia	24
3.3	Powerset	26
3.4	Twine	28
3.5	THESEUS – Ein Projekt der Bundesregierung	30
3.6	Bing	33
3.7	Wolfram Alpha	35
3.8	eyePlover	37
3.9	Qimaya	39

4.	Semantic Web auf dem Weg in den Massenmarkt	42
5.	Kritik und Machbarkeit	44
6.	Fazit	46
7.	Literaturverzeichnis	49
7.1	Bücher	49
7.2	Zeitschriften	49
7.3	Zeitungen	49
7.4	Onlinequellen	50
7.5	Persönliche Gespräche	53
	Selbständigkeitserklärung	54

ABBILDUNGSVERZEICHNIS

Abbildung 1: Funktionsprinzip einer Ontologie	15
Abbildung 2: RDF-Tripel	17
Abbildung 3: simpler RDF-Graph	18
Abbildung 4: Suchmaschine Evri mit den Suchergebnissen zu „Angela Merkel“	22
Abbildung 5: Suchmaschine Hakia mit den Suchergebnissen zur Frage „What is the meaning of Charité?“	24
Abbildung 6: Suchmaschine Bing mit den Ergebnissen zu „Amazon“	33
Abbildung 7: Suchmaschine Wolfram Alpha und die Ergebnisse zur Frage „What is the time in New York?“	35
Abbildung 8: visuelle Suchergebnisse beim eyePlover	38
Abbildung 9: Qimaya und die Ergebnisliste zum Suchbegriff „Hund“ ..	40

ABKÜRZUNGSVERZEICHNIS

CSS:	Cascading Style Sheets
HTML:	HyperText Markup Language
OWL:	Web Ontology Language
QDEX:	Query Detection and Extraction
RDF:	Resource Description Framework
RDFS:	Resource Description Framework Schema
SEO:	Search Engine Optimization
URL:	Uniform Resource Locator
URI:	Uniform Resource Identifier
W3C:	World Wide Web Consortium
XHTML:	Extensible HyperText Markup Language
XML:	Extensible Markup Language

1. EINLEITUNG

Was vor rund 40 Jahren noch zur Vernetzung von Universitäten und Forschungseinrichtungen gedient hat, ist mittlerweile zum absoluten Massenmedium für jedermann geworden - das Internet. Nach Schätzungen der BITKOM¹ werden 2010 ca. 1,5 Milliarden Menschen das Internet nutzen². Dabei dient es auch seit Facebook, StudiVZ und Co. eben nicht primär der Unterhaltung und Kommunikation, sondern immer noch hauptsächlich der Informationsbeschaffung. Informationen gibt es dabei wie Sand am Meer, doch wird es zunehmend schwerer bei den mittlerweile mehreren Milliarden Seiten wirklich relevante Informationen zu finden.

Einer der ersten, der sich mit dem Problem der Informationslogistik beschäftigte, war Vannevar Bush. Er erkannte schon in den 30er Jahren das Problem der Informationsarchivierung und -beschaffung und erfand den Memex³, ein analoges Gerät zur Speicherung von wissenschaftlichen Artikeln auf Mikrofilm, deren Wiederfindens und zu deren Verwaltung. Im Juli 1945 erschien im „The Atlantic Monthly“ Bushs Artikel „As we may think“, in dem er, passend zum Ende des zweiten Weltkrieges, den Ruf nach einer neuen Beziehung zwischen denkenden Menschen und der Summe all unseres Wissens äußerte. Er vertrat in seinem rhetorisch mitreißenden Artikel nicht nur die Meinung, dass endlich ein Gerät dafür erfunden werden müsste, weil es den Menschheits Traum der Bündelung alles menschlichen Wissens real werden lassen könnte, sondern auch, weil er meinte, zu seiner Zeit stünden die grundsätzlichen Mittel zur Verfügung, dieses Gerät auch technisch umzusetzen und einer breiten Masse zur Verfügung zu stellen. Er beschreibt seine Ideen bis ins Detail der ideellen und technischen Realisierung und schlägt bei Problemstellungen die gleichen Lösungen vor, die weiterhin das Ein-mal-eins des Informationszeitalters darstellen.

¹ Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.

² Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. , Fast jeder fünfte Mensch ist online, URL: http://www.bitkom.org/de/presse/49919_46069.aspx, Abruf: 26.08.2009

³ Akronym für „memory extender“

Ein ganz wichtiger Punkt, den Bush damals auch schon kritisierte, war die Indizierung und Selektion der Information nach Argumenten. Er äußert seine Meinung dazu wie folgt:

“Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it. In minor ways he may even improve, for his records have relative permanency. The first idea, however, to be drawn from the analogy concerns selection. Selection by association, rather than indexing, may yet be mechanized. One cannot hope thus to equal the speed and flexibility with which the mind follows an associative trail, but it should be possible to beat the mind decisively in regard to the permanence and clarity of the items resurrected from storage.”⁴

Dieses letzte große Problem und die Tatsache, dass Bush ein begeisterter technischer Visionär, aber weniger ein handwerklicher „Daniel Düsentrieb“ war, führten dazu, dass es bei der Veröffentlichung des Artikels blieb und die Maschine nie gebaut wurde.

Heute haben wir mit dem PC die notwendige Maschine und mit dem Internet die passende Technologie - seine Vision wurde Wirklichkeit. Doch das Problem der Indizierung und Selektion von Informationen nach Argumenten besteht zu großen Teilen nach wie vor. Abhilfe könnte jetzt ein Konzept schaffen, das auf einem Vorschlag des Internetgründers Tim Berner-Lee beruht - das Semantic Web.

⁴ Bush 1945, As We May Think, URL: <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml>, Abruf: 26.08.2009

2. DIE IDEE DES SEMANTIC WEB

In diesem Abschnitt werden die grundlegenden Ideen dessen, was das Semantic Web ausmacht dargestellt und die dafür zentralen Begriffe erläutert.

2.1 PROBLEME DES INTERNETS

Das Internet präsentiert sich von Tag zu Tag mehr mit einer unüberschaubaren Menge präsenter Informationen, deren Repräsentation auf den Menschen als Endnutzer ausgerichtet ist. Und genau da liegt das Problem. Der Mensch als Benutzer kann problemlos die Informationen auf Webseiten erfassen, in andere Darstellungsformen transformieren und zu anderen Informationen in Beziehung setzen. Eine Maschine hingegen kann dies in der Regel noch nicht leisten. Das Problem besteht nun darin, dass bestimmte Informationen, die zwar theoretisch verfügbar sind, in der Masse der vorhandenen Informationen im World Wide Web nur sehr schwer gefunden werden können. Aktuelle Suchmaschinen wie Google und Co. leisten hier durch den Einsatz ausgefeilter statistischer Methoden zwar schon Erstaunliches, diese Technologien jedoch basieren nur auf der Lokalisierung von Zeichenketten im Text. Besser wären hier weniger stichwortbasierte, mehr inhaltliche, eben semantische Suchmaschinen. Einer Suche also, die die Bedeutung von Inhalten interpretieren und verarbeiten kann.

Weitere Probleme ergeben sich durch die dezentrale Struktur und Organisation des Internets, die zwangsläufig zu einer Heterogenität der vorhandenen Informationen führen. Unterschiedliche Dateiformate bzw. Kodierungstechniken, die verschiedenen Sprachen bis hin zum unterschiedlichen Aufbau vieler Internetseiten machen es fast unmöglich über das Web verteilte Informationen zu bündeln, einheitlich aufzubereiten und weiterzuverwenden. Hier haben wir es mit einem Problem der Informationsintegration zu tun. Schließlich wäre es denkbar, dass ein Nutzer nach einer Information sucht, die so vielleicht gar nicht im Internet zu finden ist, sondern lediglich aus einer Reihe verschiedener

Fakten, die im Web verteilt sind, zu schlussfolgern ist. Beispielsweise ergibt die Information, dass San Fernando in Venezuela und Berlin in Deutschland liegt, gemeinsam mit der Information, dass die Differenz der lokalen Zeiten fünf Stunden beträgt, dass man aus Berlin möglichst nicht vor der Mittagszeit in San Fernando anrufen sollte. Hierbei handelt es sich um die Problematik des impliziten Wissens.

2.2 DIE LÖSUNG – DAS SEMANTIC WEB

Zur Lösung dieser Probleme kann man grundsätzlich zwei komplementäre Herangehensweisen unterscheiden.

Einerseits wäre es denkbar Methoden der Künstlichen Intelligenz anzuwenden, die die kognitiven Aufgaben, die ein Mensch bei der Verarbeitung von Informationen erfüllen würde, übernehmen und die primär für den Mensch repräsentierten Daten in maschinenlesbare Daten umwandelt.

Eine alternative Herangehensweise besteht im Ansatz des Semantic Web. Das Semantic Web steht dabei für die Idee, die Informationen schon von vornherein so zur Verfügung zu stellen, dass sie für Maschinen lesbar und vor allem interpretierbar sind. Während also das World Wide Web eine Möglichkeit darstellt alle Daten der Welt miteinander zu vernetzen, zeigt das Semantic Web einen Weg auf, um die Informationen der Welt auf der Ebene ihrer Bedeutung miteinander zu verknüpfen.

Dafür ist es zwingend erforderlich einheitliche und offene Standards für die Beschreibung von Informationen zu vereinbaren, die es letztendlich ermöglichen sollen, Informationen verschiedener Anwendungen und Plattformen auszutauschen und zueinander in Beziehung zu setzen. Diese Fähigkeit wird auch Interoperabilität genannt. Dabei ist es nicht nur wichtig, dass die Standards formal klar definiert sind, sondern dass sie auch flexibel und erweiterbar sind. Das heißt, dass sie auch für unvorhergesehene Anwendungsfälle funktionieren müssen ohne

dabei später einmal revidiert werden zu müssen. Die Ausarbeitung und Beschreibung solcher Standards hat sich das „World Wide Web Consortium“⁵ zur Aufgabe gemacht. Gegründet vom Internet-Erfinder Tim Berner-Lee, definiert das W3C seit 1994 neben Standards zu HTML, XML, XHTML, CSS und weiteren für das Internet standardisierte Techniken, auch grundlegende Standards für das „Resource-Description-Framework“⁶ und die „Web Ontology Language“⁷. Letztere sind Informations-Spezifikationssprachen bzw. Ontologiesprachen, die speziell für die Verwendung im Semantic Web entwickelt wurden.

Ein weiterer wichtiger Punkt für die Umsetzung des Semantic Web sind Methoden zur Schlussfolgerung von „neuen“ Informationen aus bereits gegebenen. Um größtmöglichen Nutzen aus spezifiziertem Wissen zu ziehen, muss auch implizite Information aus einer Spezifikation extrahiert werden können.

Im weiteren Verlauf meiner Arbeit schreibe ich oftmals davon, Maschinen würden die Bedeutung von Informationen verstehen. Dabei sei erwähnt, dass das nicht das grundlegende Ziel des Semantic Web darstellt. Diese sind wesentlich moderater. Es sollen viel mehr Wege und Methoden gefunden werden, *„Informationen so zu repräsentieren, dass Maschinen damit in einer Art und Weise umgehen können, die aus menschlicher Sicht nützlich und sinnvoll erscheint. Dabei mag es das entstehende Verhalten von Systemen in den Augen mancher Nutzer durchaus als intelligent erscheinen.“*⁸

⁵ nachfolgend W3C genannt

⁶ nachfolgend RDF genannt

⁷ nachfolgend OWL genannt

⁸ Hitzler/ Krötzsch / Rudolph et al. 2008, 12

2.3 DIE BEGRIFFE SEMANTIK UND SYNTAX

Zwei Begriffe, die bei der Definition von Sprachen und in der Logik eine zentrale Rolle spielen, sollen hier kurz definiert werden.

„Unter Syntax versteht man im Allgemeinen eine Menge von Regeln zur Strukturierung von Zeichen und Zeichenketten. In der Linguistik steht Syntax für eine Menge von Regeln, um aus Worten Sätze oder Phrasen zu bilden. In der Informatik, und auch im Sinne [dieser Arbeit] [...], steht Syntax für eine Menge von Regeln, um Programme oder Dokumente mit bestimmten Eigenschaften, z.B. gültige XML-Dokumente, zu erzeugen. Die Semantik steht allgemein für ‚Bedeutung von Wörtern, Phrasen oder Symbolen‘. In der Linguistik steht Semantik für die Wissenschaft der Bedeutung in verschiedenen Sprachformen. In der Informatik, und insbesondere im Bereich Semantic Web versteht man unter Semantik die Bedeutung von Worten bzw. Zeichen(-ketten) und ihre Beziehungen untereinander.“⁹

Allgemein kann man also sagen, dass im Sinne dieser Arbeit Syntax für die normative Struktur der Daten steht, welche erst durch die Semantik eine Bedeutung erhält.

⁹ Hitzler/ Krötzsch / Rudolph et al. 2008, 13

2.4 ANNOTATIONEN

Im Unterschied zum Information-Retrieval¹⁰ mit Informationsextraktion, das auf unstrukturierten Daten aufbaut, setzt das Semantische Web Annotationen (Metadaten) voraus. Die Bedeutung von Informationen werden also bei der Programmierung von Internetseiten mit Hilfe einer Auszeichnungssprache explizit dazu geschrieben.

Annotationen sind also nichts anderes als Mittel zur Strukturierung von Programmquelltexten. Die Annotation geschieht unter Einsatz von festgelegten Vokabularien und Ontologien, beispielsweise mit RDF oder OWL. Auf letztere werde ich etwas später noch genauer eingehen. Einerseits können durch Annotationen Internetseiten besser kategorisiert werden, zum Anderen wird es ermöglicht Schlussfolgerungen zu treffen.

Beispielsweise besagt die Annotation einer Internetseite, dass sie sich mit Thema „100m-Lauf“ beschäftigt. Aus der verwendeten Ontologie würde dann hervorgehen, dass das übergeordnete Thema „Leichtathletik“ ist und dass die „Leichtathletik“ eine Sammelbezeichnung für verschiedene Sportarten darstellt obwohl dies nicht ausdrücklich in den Metadaten hinterlegt wurde.

Durch eine entsprechend gute Begriffswahl in der Annotation könnte dadurch ein hoher Automatisierungsgrad bei der Verarbeitung von Internetseiten erzielt werden. Die Annotation stellt somit einen sehr wichtigen Punkt dar, wenn es darum geht, dass Suchmaschinen uns zukünftig komplexe Fragen beantworten. Das Suchergebnis der Frage „Welche Zeit brachte Usain Bolt bei seinem 100m-Lauf der Leichtathletik-WM 2009 einen erneuten Weltrekord ein?“ könnte dann lediglich diese eine benötigte Information enthalten - nämlich „9,58 Sekunden“.

¹⁰ Fachgebiet, das sich mit computergestütztem, inhaltsorientiertem Suchen beschäftigt

2.5 ONTOLOGIEN

Zur Darstellung komplexer Wissensbeziehungen wird im Fachbereich Informatik der Begriff **Ontologie** verwendet. Im Gegensatz zur Taxonomie, die einfache Hierarchien verwendet, verkörpert die Ontologie ein Netz von Hierarchien, in dem Informationen durch logische Beziehungen miteinander verknüpft sind oder sein könnten. Diese Beziehungen beruhen auf Eigenschaften, die den Informationen spezifisch zugewiesen werden müssen. Elemente, die auf diese Weise zusammenhängen, sind dann semantisch erzeugt. Ontologien bestehen aus einer Vielzahl von Komponenten wie beispielsweise Begriff (Konzept), Instanzen und Relationen.

Abbildung 1: Funktionsprinzip einer Ontologie

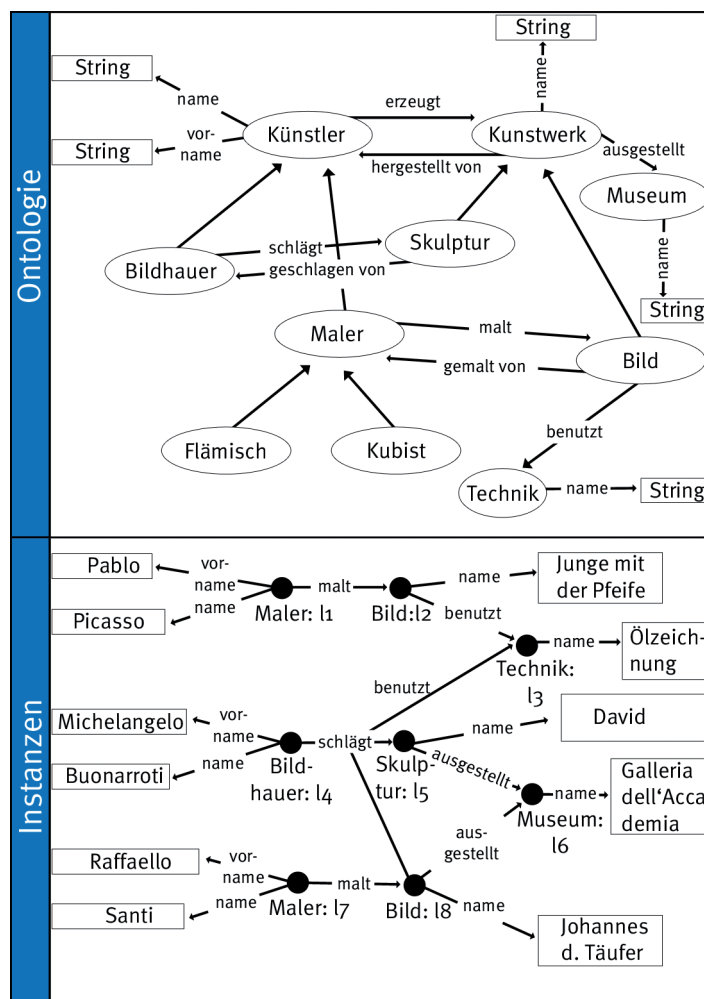


Abbildung 1 zeigt das Funktionsprinzip einer Ontologie. Der obere Teil zeigt dabei die Ontologie, die Begriffe und Relationen enthält, wobei Begriffe durch Ellipsen und Relationen durch Pfeile gekennzeichnet sind. Die Rechtecke stellen einfache Container für Informationen dar. Die Relationen verbinden zwei Begriffe miteinander und schränken diese gleichzeitig ein, beispielsweise wird ein Kunstwerk von einem Künstler erzeugt.

Begriffe können zur Vererbung herangezogen werden. Aus diesem Grund besitzen die Maler und Bildhauer ebenfalls die Relationen Name und Vorname. Der dicke Pfeil kennzeichnet die Vererbung. Die beiden Relationen „schlägt“ und „malt“ sowie „gemalt von“ und „geschlagen von“ sind vererbte Relationen von „erzeugt“ und „hergestellt von“. Die ursprünglichen Relationeigenschaften bleiben dabei erhalten, können jedoch erweitert werden.

Die Relationen „malt“ und „gemalt von“ besitzen inverse Beziehungen zueinander, wodurch weitere Logik in die Ontologie integriert wird, die es ermöglicht, dass von einem Maler auf seine Kunstwerke und umgekehrt, von einem Bild zum Maler, geschlossen werden kann.

Der untere Teil der Abbildung zeigt die Instanzen der Ontologie. Diese werden durch einen schwarzen Punkt dargestellt. Das Kürzel (I₁) steht dabei für den einmaligen Ressourcennamen der Instanz. Im Semantischen Web wird ein „Uniform Resource Identifier“¹¹ zur Kennzeichnung verwendet.

Der URI dient zur Identifizierung einer abstrakten oder physischen Ressource. Er stellt also sicher, dass es zu keinen Verwechslungen zwischen Objekten und/oder Relationen kommt. Semantische Eindeutigkeit wird dadurch möglich. Die im Allgemeinen geläufigere Abkürzung URL (Uniform Resource Locator) stellt dabei eine Unterart des URI dar und gibt lediglich den Ort einer Ressource im Netz an.

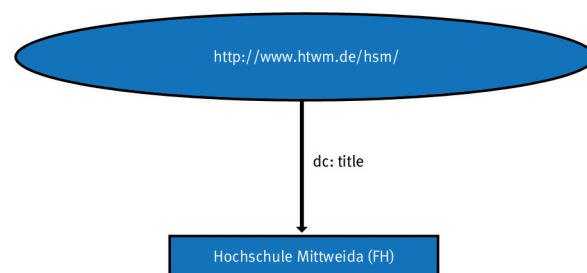
¹¹ nachfolgend URI genannt

Zusammenfassend kann man sagen, dass Ontologien Modelle sind, die vergleichbar mit relationalen Datenbanken sind, dabei jedoch weitaus flexibler. Sie geben eine Ordnungsstruktur vor, die es Maschinen möglich macht, sinnvolle Zusammenhänge zu erkennen und den Kontext zu erkennen, aus dem sich der Nutzer einem Thema nähert. Ontologien verwalten sozusagen die Metadaten und ihre Relationen.

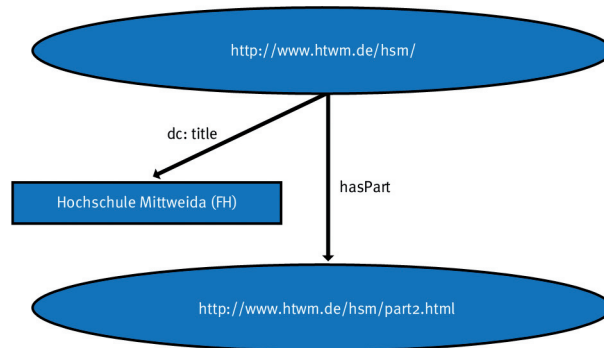
2.6 RDF/XML

Das RDF (Resource Description Framework) ist eine Auszeichnungssprache für Metadaten und basiert auf so genannten „triples“ oder „statements“, die wiederum aus „subject“, „predicate“ (oder auch „property“) und „object“ bestehen. Diese Tripel kann man als Erweiterung zu Schlüssel-Wert-Paaren sehen. Während Schlüssel-Wert-Paare allerdings nur einer beliebigen Eigenschaft genau einen beliebigen Wert zuordnen kann (z.B. Kontaktadresse = Musterstraße), kann mit einem Tripel auf semantische Art ein Objekt, Konzept oder Wert mit einem anderen in Beziehung gesetzt werden. Ein Beispiel für ein Tripel wäre der Satz: Musterstraße ist Kontaktadresse von Max Mustermann. Bei diesem Satz wäre „Musterstraße“ das Subjekt, „ist Kontaktadresse von“ das Prädikat und „Max Mustermann“ das Objekt.

Abbildung 2: RDF-Tripel



Ein weiteres Beispiel zeigt die Abbildung 2. Die Grafik zeigt ein RDF-Tripel und lässt sich folgendermaßen interpretieren: „Hochschule Mittweida (FH)“ ist der Titel von „`http://www.htwm.de/hsm/`“.

Abbildung 3: simpler RDF-Graph

Mehrere RDF-Tripel bilden dann wiederum einen RDF-Graphen, mehrere RDF-Graphen ergeben dann die Metadaten als solche.

Da idealerweise ein für die Eigenschaften bekanntes und weitverbreitetes Vokabular benutzt wird, wie z.B. das Dublin Core Element Set (DC)¹², das eindeutige URIs für die wichtigsten Metadatentypen bereitstellt, sind die Informationen der so ausgezeichneten Ressourcen auch für Computerprogramme als Metadaten identifizierbar und entsprechend interpretierbar, also z.B. ein Autor als eben solcher. Die RDF-Daten können in dem vom W3C spezifizierten XML-Format RDF/XML abgebildet werden. Sie können aber ebenso auch in für Menschen lesbaren Formaten wie „Notation 3“ (auch N3 oder RDF/N3 genannt) repräsentiert werden.

¹² Dublin Core ist eine Sammlung einfacher und standardisierter Konventionen zur Beschreibung von Dokumenten und anderen Objekten im Internet, um diese mit Hilfe von Metadaten einfacher auffindbar zu machen

2.7 RDF-SCHEMA

Mit dem RDF-Modell erhält man die Möglichkeit XML-konforme Dokumente zu erzeugen, welche die Objekte anhand von Tripel beschreiben. Möchte man hingegen eine Gruppe von ähnlichen Objekten mit den gleichen Eigenschaften auszeichnen, bietet das RDF leider keine Möglichkeit. Dafür wurde das RDF-Schema¹³, welches offiziell „RDF Vocabulary Description Language“ genannt wird, definiert. Mit diesem Schema hat man die Möglichkeit Begriffe und die damit verbundenen Elemente semantisch zueinander in Beziehung zu setzen. In RDFS wird für jede Eigenschaft festgelegt, welche Werte erlaubt sind, was diese für eine Bedeutung haben, welche Beziehungen zu anderen Eigenschaften bestehen und welche Arten von Ressourcen diese Eigenschaft verwenden dürfen. Vom W3C wurde dafür allerdings kein allgemein gültiges Schema definiert, in dem verschiedene Klassen und Eigenschaften festgelegt werden. Stattdessen wird es in einer „Scheme-Definition Language“ beschrieben, mit deren Hilfe die eigentlichen Schemata definiert werden können. Die Schemata werden auch als Vokabulare bezeichnet. In den letzten Jahren haben sich RDFS-Gemeinschaften, wie z.B. der Dublin Core gebildet, die die Aufgabe haben, RDFS-Metadatenmodelle zu entwerfen, da man sich eingestanden hat, dass es unmöglich ist, ein einzelnes Schema zu entwickeln, das für alle Gebrauchsmöglichkeiten passend wäre.

¹³ nachfolgend RDFS genannt

2.8 WEB-ONTOLOGIESPRACHE (OWL)

Die „Web Ontology Language“¹⁴ wurde vom W3C erarbeitet und standardisiert und zählt daher aktuell als die populärste Sprache zur Modellierung von Ontologien. Mit OWL werden, genau wie mit RDFS, Terme einer Domäne und deren Beziehungen formal beschrieben. OWL hat im Vergleich zu RDFS aber weitaus komplexere Funktionen zum Beschreiben der Beziehungen. Des Weiteren lassen sich Konzepte in OWL deutlicher spezifizieren, wodurch ein höherer Abstraktionsgrad entsteht. Zusätzlich zu RDF und RDFS werden in der OWL weitere Sprachkonstrukte eingeführt, die es erlauben Ausdrücke ähnlich der Prädikatlogik zu formulieren. Dadurch wiederum können dann bessere logische Schlussfolgerungen gezogen werden.

Mit RDF und OWL sind die technologischen Bausteine für das Semantic Web also bereits vorhanden und halten sukzessive Einzug in unterschiedlichste Anwendungsbereiche.

¹⁴ nachfolgend OWL genannt

3. BEISPIELE VON SEMANTISCHEN WEB-ANWENDUNGEN ODER PROJEKTEN MIT BEZUG ZUM SEMANTISCHEN WEB

Nachdem die grundlegenden Konzepte und Technologien des Semantic Web genannt sind, sollen auf den folgenden Seiten nun einige ausgewählte Beispiele von derzeit schon existierenden semantischen Webanwendungen dargestellt und etwas näher erklärt werden. Dabei soll nicht ausschließlich auf semantische Suchmaschinen wie Bing und Co. eingegangen werden, sondern auch auf andere semantische Projekte wie beispielsweise dem Web-Organizer Twine, oder THESEUS, einem Projekt der Bundesregierung.

3.1 EVRI

„Search Less, understand more“¹⁵ ist das Motto des amerikanischen Startup Unternehmens Evri. Dabei stellt Evri keine herkömmliche Suchmaschine dar, wie man anfangs erwarten könnte. Viel mehr versteht sich Evri als Web-Service, der Hilfestellung zum Auffinden relevanter und miteinander in Beziehung stehender Informationen anbietet. Ziel soll es dabei sein, dem Nutzer mehr relevante Inhalte in kürzerer Zeit zu liefern. Dabei wird eine semantische Verknüpfung benutzt, die Inhalte wie News-Artikel, Bilder und Videos einem Suchbegriff zuordnet und als Aggregat darstellt. Evri stellt die Ergebnisse der Suchanfrage nicht auf die herkömmliche Art und Weise in Form einer Auflistung mehrerer relevanter Links zum eingegebenen Suchbegriff dar, sondern erstellt über spezielle Profilseiten ein einfaches Beziehungsschaubild.

¹⁵ <http://www.evri.com/>

Abbildung 4: Suchmaschine Evri mit den Suchergebnissen zu „Angela Merkel“

The screenshot shows the Evri search engine interface. At the top, there's a search bar with the text 'Go to:' and a search button. Below the search bar, the results for 'Angela Merkel' are displayed. On the left, there's a profile card for Angela Merkel, including a photo, her name, and a brief description. To the right of the profile card, there's a 'Recent connections' section showing a network of related entities: Dmitry Medvedev, CDU, Germany, Magna International, and Opel. Below this, there's a 'More on Angela Merkel' section with various links and tags. On the right side of the page, there's a list of articles, images, quotes, and videos related to Angela Merkel. The 'Articles' section includes a headline about Medvedev and Merkel condemning Chechnya killings. The 'Images' section shows four small images. The 'Quotes' section features a quote from August 07, 2009. The 'Videos' section includes a video titled 'WMX 61 "Vlogging with German Chancellor Angela Merkel"'. At the bottom, there's a 'Trends' section showing a bar chart of mentions in the last 30 days.

Die Abbildung zeigt die Ergebnisse für den Suchbegriff „Angela Merkel“ und dazu die fünf Begriffe, die laut Evri den größten Bezug zum gesuchten Wort haben: „Dmitri Medvedev“, „CDU“, „Germany“, „Magna International“ und „Opel“. Die von Evri in Verbindung gebrachten Begriffe stellen sich allesamt als sehr brauchbar und vor allem aktuell dar. Denn zwei Tage vor Aufruf dieser Seite war Angela Merkel auf einem Treffen mit dem Präsidenten der russischen Föderation Dmitri Medvedev und ca. zwei Wochen zuvor plädierte die Bundeskanzlerin für eine Übernahme des Opel-Konzerns durch den Autozulieferer Magna International.

Dies nur am Rande, zur Deutlichmachung der Aktualität und Relevanz der von Evri in Verbindung gebrachten Begriffe. Neben dem Schaubild der fünf in Beziehung gebrachten Begriffe werden einige Top-Artikel zum Begriff gelistet und es werden Bilder, sowie Videos in einer Vorschau angezeigt. Außerdem wird der Begriff mit einer Kurzbeschreibung aus der Wikipedia-Datenbank verknüpft. Durch einen einfachen Klick lässt sich der ursprüngliche Suchbegriff mit einem der dazugehörigen Top-Begriffe verknüpfen, um so Artikel, Bilder und Videos mit beiden Begriffen zu finden.

Um eine Anfrage bei Evri starten zu können, muss der gesuchte Begriff als Profil bereits definiert sein. Allerdings baut Evri nicht nur auf die Weiterentwicklung des eigenen Web-Service, sondern bietet auch Widgets an, die auf anderen Internetseiten integriert werden können. Dort wird dem Leser angezeigt, mit welchen anderen Inhalten die Personen, der Ort, oder die News in dem Artikel verknüpft sind. Damit soll es dann ermöglicht werden alle relevanten Informationen zu einem Thema zu erhalten, über das man gerade etwas liest.

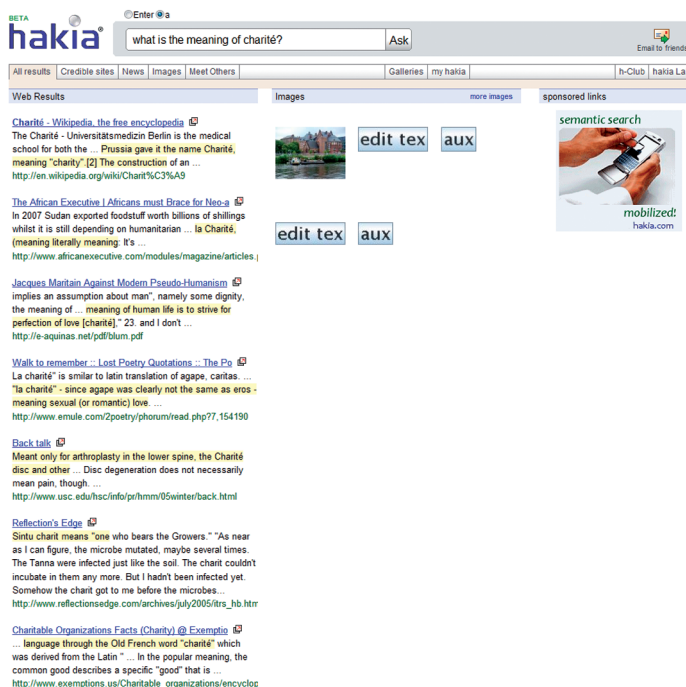
Die Funktionalität vom Web-Service Evri basiert nicht auf Algorithmen, wie sie bei vielen Suchmaschinen verwendet werden, sondern identifiziert jede Person, jeden Ort und jeden Begriff in einem Artikel und untersucht dabei deren Beziehungen zueinander. Diese Informationen stellen die Basis zum Aufzeigen anderer themenspezifischer Inhalte dar.

Evri stellt momentan eine sehr brauchbare Alternative zu Google dar, soll aber auch in Zukunft keine herkömmliche Suchmaschine ersetzen, sondern lediglich als Unterstützung dienen, wenn es darum geht brauchbare Inhalte im Internet zu finden.

3.2 HAKIA

Bei Hakia¹⁶ handelt es sich um eine semantische Suchmaschine. Die Macher von Hakia fordern regelrecht dazu auf ausformulierte Fragen und nicht wie sonst üblich nur die gesuchten Begriffe einzugeben, was allerdings trotzdem funktioniert. Bei der Suche werden nicht nur genau die Worte gesucht, die der Benutzer übermittelt hat. Vielmehr wird versucht die Bedeutung der eingegeben Wörter zu erkennen und Ergebnisse mit einzubeziehen, die der Nutzer mit seiner Wortwahl eventuell gar nicht berücksichtigt hat.

Abbildung 5: Suchmaschine Hakia mit den Suchergebnissen zur Frage „What is the the meaning of Charité?“



Das, was man zunächst nicht wirklich glauben kann, funktioniert schon beeindruckend gut. Hakia erkennt ausformulierte Fragen und, was viel wichtiger ist, es werden sinnvolle und glaubwürdige Antworten ausgegeben. Auf Abbildung 5 sieht man die Suchergebnisse der Frage „What is the meaning of Charité?“. Der erste Treffer liefert bei dieser Suchanfrage schon das richtige Ergebnis.

¹⁶ <http://www.hakia.com/>

Die Antwort-Seite ist jedes Mal in drei Spalten unterteilt. In der linken Spalte werden die Ergebnisse angezeigt, in der mittleren Spalte findet man Ergebnisse von ausdrücklich vertrauensvollen Seiten, sowie einige Bilder, News oder „User generated content“¹⁷. Rechts bleibt Platz für Anzeigen, Banner oder Werbung, die zur Finanzierung der Webseite dienen. In den angezeigten Ergebnissen selbst wird die vermeintliche Antwort durch eine gelbe Markierung hervorgehoben.

Technisch realisiert wird das Ganze durch das eigens für Hakia entwickelte Query Detection and Extraction¹⁸ System. QDEX ist eine neue Art und Weise Internetseiten zu analysieren. Beim Durchsuchen einer Internetseite extrahiert der QDEX-Algorithmus alle möglichen Fragen die zum Inhalt aufkommen könnten und speichert diese. Die meisten Suchmaschinen arbeiten mit einem invertierten Index, der ein riesiges aktives Daten-Set beinhaltet. Würde man diesen invertierten Index mit semantischen Informationen anreichern, würde die operative Belastung der Suchanfrage auf exponentielle Weise ansteigen, was die Dauer einer Suchanfrage erheblich erhöhen würde. Daher benutzt das QDEX-System kleine Daten-Sets für jede Suchanfrage, was die Abarbeitung der semantischen Assoziationen wesentlich schneller macht.

Der kritische Punkt beim QDEX-System ist es, Sätze in wenige sinnvolle Sequenzen zu zerlegen. Ein Satz, der beispielsweise acht bedeutungsvolle Wörter enthält, könnte mehr als eine Milliarde Sequenzen generieren, von denen gerade einmal ein Dutzend sinnvoll wären. Aufgabe ist nun aus diesen Milliarden Sequenzen die sinnvollen heraus zu filtern. Um das zu ermöglichen benutzt Hakia die OntoSem-Technologie.

¹⁷ steht für Inhalte, die nicht vom Anbieter eines Webangebots, sondern von dessen Nutzern erstellt werden

¹⁸ nachfolgend QDEX genannt

3.3 POWERSET

Im Mai 2008 startete offiziell der nächste Angriffsversuch auf Googles Vorherrschaft im Suchmaschinengeschäft. Dieses Mal versuchte das in San Francisco ansässige Unternehmen Powerset¹⁹ mit seiner gleichnamigen Suche die Suchmaschine neu zu erfinden. Wie das gehen soll, verrät Powersets Chief Operating Officer (COO) Steve Newcomb: „*Wir erfinden die Internet-Suchmaschine neu*“²⁰, sagt er über die Technologie, mit der Powerset seinen Angriff Mitte des Jahres starten will. „*Höchste Zeit, dass jemand den Mumm dazu hat.*“²¹

Powerset versteht natürliche Sprache und kann Antworten auf ausformulierte Fragen geben. Gibt der Nutzer einen Satz oder eine Frage ein, analysiert Powerset diesen Ausdruck und erkennt dessen Bedeutung. Dazu wird eine Technik zur Verarbeitung natürlicher Sprache verwendet, die am legendären XEROX-Entwicklungspark PARC (Palo Alto Research Center) entwickelt wurde, das bei der Entwicklung von Natural-Language-Anwendungen an vorderster Front steht. „Die Suchmaschinen von heute sind funktionell beschränkt“, weiß PARC-Direktor Mark Bernstein. „*Natural Language ist die Zukunft.*“²² Oder bald schon die Gegenwart: „*Diese Technologien*“, sagt Newcomb, „*sind jetzt endlich kommerziell reif.*“²³

Anders als beispielsweise bei Google und Yahoo indexiert Powerset nicht das gesamte Web, sondern greift derzeit nur auf die Datenbestände von Wikipedia und der offenen Datenbank Freebase zu. Die Quellen sollen allerdings nach und nach erweitert werden.

¹⁹ <http://www.powerset.com>

^{20, 21, 22, 23} Manager Magazin, Neue Suchmaschine – Griff nach Googles heiligem Gral, URL: <http://www.manager-magazin.de/it/artikel/o,2828,471830,00.html> Abruf: 19.08.2009

Schon vor dem Start des neuen Suchdienstes galt Powerset als eines der interessantesten Startups in den USA. Nachdem einige Gerüchte um die Übernahme des Unternehmens kursierten, wurde Anfang Juli 2008 bekannt gegeben, dass Microsoft das seit 2005 existierende Unternehmen Powerset für 100 Millionen US-Dollar kaufen wird. Aus Sicht des Analysten Nick Patience von The 451 Group ist Microsoft allerdings weniger an der Technik von Powerset, sondern vor allem an dessen Entwicklermannschaft interessiert. *„Microsoft kauft hier eine Firma, die zwar ziemlich innovative Arbeit abgeliefert hat, aber damit noch in den Kinderschuhen steckt“*²⁴, sagt der Experte. *„Sie haben sie wegen der Leute erworben.“*²⁵

Ob sich Powerset nun mit der Unterstützung von Microsoft tatsächlich gegen Google durchsetzen kann bleibt allerdings weiter abzuwarten, denn vorerst müssen die Entwickler beweisen, dass sich das, was bisher beim Datenbestand von Wikipedia schon sehr gut funktioniert auch auf das gesamte Web übertragen lässt.

^{24, 25} Computerwoche, Microsoft kauf Powerset, URL: <http://www.computerwoche.de/netzwerke/web/1867840/> Abruf: 19.08.2009

3.4 TWINE

Im Oktober 2008 wurde der Web-Organizer Twine²⁶ gestartet, der mit speziellen semantischen Ansätzen arbeitet. Entwickelt wurde der Online-Dienst vom Startup „Radar Networks“ aus San Francisco.

Twine stellt eine Mischung aus Lesezeichen-Werkzeug, sozialem Netz und Empfehlungsmaschine dar und hilft den Nutzern dabei Informationen aus dem Internet zu sammeln, sie zu managen und sie mit anderen Nutzern zu teilen. Durch den sehr hohen Funktionsumfang, den der Dienst bietet, wendet sich Twine vor allem an fortgeschrittene Benutzer. Für Einsteiger sind die vielen Funktionen zunächst nicht ganz leicht erfassbar.

Nachdem man sich bei Twine einen Account erstellt hat, kann man sich ein so genanntes Bookmarklet in die Leiste seines Browsers ziehen. Damit kann man Internetseiten, die man sich merken will, mit einem Klick automatisch bei Twine erfassen. Des Weiteren lassen sich natürlich auch die gespeicherten Lesezeichen aus Firefox und Co., sowie anderen Lesezeichen-Diensten problemlos übertragen.

Auf den ersten Blick hat das Ganze nicht viel mit Semantik zu tun, viel mehr wirkt es wie andere Social Networking-Anwendungen auch. Nutzer stellen Kontakte zu anderen Nutzern her, tauschen Informationen aus und diskutieren darüber. Künstliche Intelligenz, maschinelles Lernen und die Verarbeitung natürlicher Sprache werden dem Nutzer auf den ersten Blick nicht bewusst. „*Das semantische Web ist eine nützliche Technologie. Es ist aber ein Mittel zum Zweck und nicht der Zweck an sich*“, sagt Nova Spivack, der Chef von Radar Networks. „*Wir wollen den Nutzern mit diesem Release und auch in Zukunft stets zeigen, was man mit Twine machen kann. Dass da das semantische Web dahinter steckt, ist nur ein Detail für Geeks.*“²⁷

²⁶<http://www.twine.com/>

²⁷ Heise Online, Organizer für das semantische Web, 03.11.2008,
URL: <http://www.heise.de/tr/Organizer-fuer-das-semantische-Web-/artikel/118218>
Abruf: 14.08.2009

Twine nutzt verschiedene Verfahren aus der Künstlichen Intelligenz, Algorithmen aus dem Bereich des maschinellen Lernens und der Verarbeitung natürlicher Sprache. Mit diesen Techniken werden die Inhalte einer Web-Seite analysiert und die grundlegenden Konzepte extrahiert – Personen, Orte, Organisationen und einiges mehr. Mit dem Ergebnis lassen sich dann Verbindungen zu bereits erfassten Informationen auf anderen Seiten herstellen, auch weitere Twine-Nutzer, die ähnliche Interessensgebiete haben, werden aufgefunden.

Wenn man beispielsweise einen „Twine“ zu einer speziellen Programmiertechnik anlegt, bekommt man automatisch auch Lesezeichen aus übergeordneten Bereichen angezeigt – von anderen Programmiertechniken bis hin zu Anwendungen, die diese Programmiertechnik verwenden. Außerdem werden einem andere Nutzer vorgestellt, die gleiche oder ähnliche Interessen angegeben haben. Weiterhin kann Twine automatisch Tags, Beschreibungen und Zusammenfassungen von abgelegten Webseiten erstellen.

Derzeit arbeiten die Entwickler von Twine an weiteren Funktionen, wie beispielsweise einem Kalender, der Veranstaltung verschiedener Art intelligent gruppiert. Damit will man den derzeit noch Web 2.0 anmutenden Networking Dienst auch für die Nutzer zusehend zu einem semantischen Networking Dienst machen.

3.5 THESEUS – EIN PROJEKT DER BUNDESREGIERUNG

Auch die Bundesregierung beteiligt sich mit dem vom Bundesministerium für Wirtschaft und Technologie initiierten Forschungsprogramm THESEUS an der Entwicklung eines semantischen Internets.

Seit 2005 wird dabei das Ziel verfolgt „eine neue internetbasierte Wissensinfrastruktur zu entwickeln, um das Wissen im Internet besser zu nutzen und zu verwerten.“²⁸ THESEUS entstand aus der Quaero-Initiative, die von einer deutsch-französischen Industriearbeitsgruppe vorgeschlagen wurde.

Seit 2006 arbeiten beide Länder wieder getrennt an dem Projekt, da sich unterschiedliche Arbeitsschwerpunkte herausgebildet hatten. „THESEUS - wie der deutsche Teil seitdem heißt - und Quaero ergänzen sich komplementär und werden eng über regelmäßig tagende Arbeitsgruppen verzahnt. Es ist vorgesehen, gemeinsame Synergie- und Kooperationspotenziale zu erschließen.“²⁹ Unter dem Dach von THESEUS sollen zukünftig anwendungsorientierte Basistechnologien und technische Standards erarbeitet und erprobt werden.

Als Ergebnisse werden neuartige Produkte, Werkzeuge, Dienste und Geschäftsmodelle für das World Wide Web sowie die Dienstleistungs- und Wissensgesellschaft von morgen erwartet. Im Fokus des Forschungsprogramms stehen semantische Technologien, die Inhalte (Wörter, Bilder, Töne) nicht mit Hilfe herkömmlicher Verfahren (z. B. Buchstabenkombinationen) ermitteln, sondern die die inhaltliche Bedeutung der Informationen erkennen und einordnen können. Mit diesen Technologien können Computerprogramme intelligent nachvollziehen, in welchem Kontext Daten abgespeichert wurden.

^{28, 29} Bundesministerium für Wirtschaft und Technologie, Was ist Theseus? URL: <http://theseus-programm.de/was-ist-theseus/default.aspx> Abruf: 09.08.2009

Darüber hinaus können Computer durch Anwendung von Regeln und Ordnungsprinzipien aus den Inhalten logische Schlüsse ziehen und selbständig Zusammenhänge zwischen unterschiedlichen Informationen aus mehreren Quellen erkennen und herstellen.

Das Beispielprojekt Medico, welches aus dem Forschungsprogramm hervorging, soll künftig für eine bessere medizinische Versorgung sorgen und damit Patienten bessere Heilungschancen geben und Krankenhäusern die Möglichkeit ihre Kosten zu senken. Beispielsweise könnte Medico bei der Diagnose von schweren gesundheitlichen Problemen an den Lymphdrüsen zum Einsatz kommen. Angenommen, ein Lymphom-Patient befindet sich in einem Kreiskrankenhaus, in dem der behandelnde Arzt wissen will, ob die Chemotherapie angeschlagen hat. Mit Medico wird er nun automatisch aktuelle Bilder aus einer Computertomographie mit früheren, vor der Behandlung erstellten Bildern abgleichen können. Dabei fallen dem Programm neue Schwellungen in der Milz auf, zu denen vor Ort keine Vergleichsaufnahmen vorliegen. Digital kann der Arzt über die THESEUS-Applikation Bilder und Befunde seines Patienten an das Rechenzentrum einer großen Uniklinik schicken, um sie mit dortigen Lymphom-Fällen zu vergleichen. Auf dieser um ein Vielfaches erweiterten Informationsbasis erhält der Onkologe im Idealfall präzise Empfehlungen für die weitere Behandlung.

Die mit Medico entwickelten semantischen Techniken sollen künftig Computer dazu befähigen, selbstständig Informationen aus dem Inhalt der dargestellten Bilder zu erzeugen sowie diese mit weiteren medizinischen Daten zusammenzuführen und zu strukturieren. Dafür „*müssen noch zahlreiche Komponenten entwickelt werden*“³⁰, heißt es beim Berliner THESEUS-Projektbüro. Nötig seien etwa Verfahren zur Mustererkennung, computergestützte Erkennungssysteme oder klinische Entscheidungshilfen. Die zentrale Aufgabe bestehe in der Entwicklung eines Ordnungssystems, das die Fülle von Informationen auf Grundlage des medizinischen Fachvokabulars gliedern und vernetzen kann.

³⁰ Heise Online, Theseus soll dem Arzt helfen, URL: <http://www.heise.de/newsticker/Theseus-soll-dem-Arzt-helfen-/meldung/118793> Abruf: 12.08.2009

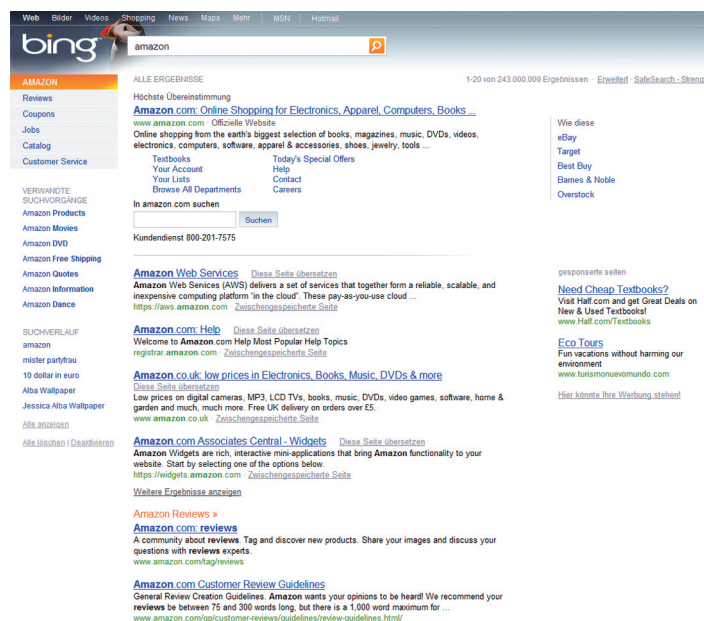
Die dafür notwendigen Metadaten sollen von Medico auf Basis der menschlichen Anatomie automatisch generiert und den jeweiligen Bilddaten zugeordnet werden. Zur Beschreibung der Bildinhalte werden Fach-Ontologien wie RadLex oder das detaillierte, frei verfügbare Foundational Model of Anatomy herangezogen.

Finanziert wird das Forschungsprogramm THESEUS mit 180 Millionen Euro, wovon 90 Millionen Euro vom Bundesministerium für Wirtschaft und Technologie kommen, die anderen 90 Millionen Euro aus der Wirtschaft. Projektführer ist die ehemalige Bertelsmann-Tochter Empolis, weitere Projekt- und Forschungspartner sind unter anderem Siemens, SAP, Bertelsmann, Fraunhofer-Gesellschaft und das Deutsche Forschungszentrum für Künstliche Intelligenz.

3.6 BING

Nach bisher zwei erfolglosen Versuchen den Suchmaschinenmarkt signifikant mit zu prägen, veröffentlichte Microsoft im Mai 2009 seinen dritten Versuch mit der Suchmaschine Bing³¹ und verfolgt dabei semantische Ansätze. Dabei versteht Bing Kombinationen aus mehreren Suchwörtern und sucht dabei nicht wie „herkömmliche Suchmaschinen“ nur Dokumente in denen eines oder mehrere der Suchbegriffe vorkommen.

Abbildung 6: Suchmaschine Bing mit den Ergebnissen zu „Amazon“



Wie auf der Abbildung zu erkennen ist, beinhaltet Bing viele kleine zusätzliche Funktionen, die den Nutzern das Gefühl geben sollen gut informiert zu sein. Beispielsweise ergibt die Suche nach „Amazon“ neben den Links zur Verkaufsplattform Amazon auch gleich die Nummer zum Kunden-Support, sowie eine Suchmaske, mit der man Amazon durchsuchen kann. Bei der Suche nach „UPS“ hingegen bekommt man neben der Nummer zum Kundendienst auch die Möglichkeit per Eingabemaske eine Paketnummer zu suchen um dieses nach zu verfolgen. Außerdem können beim Überfahren eines Links in der Ergebnisliste zusätzliche Boxen am rechten Rand

³¹ <http://www.bing.com>

eingebildet werden, in denen man weiterführende Links auf der jeweiligen Internetseite angezeigt bekommt.

Das Herzstück von Bing liegt allerdings bei der Bilder- und Videosuche. Bing wirkt hier sogar treffsicherer als Google und bietet mehr Funktionalität. Beispielsweise kann man sich per Klick auf „Hintergrundbilder“ lediglich Bilder anzeigen lassen, die der Größe der Auflösung des Nutzer-PCs entsprechen. Diese wird dabei von Bing automatisch ermittelt. Auch die Videosuche gibt einiges her. Beim Überfahren mit der Maus wird das Video in einem Vorschaufenster gezeigt, inklusive Ton. Darüber hinaus outet sich Microsoft auch als lernbegierig. Man kann Bilder nämlich auch kommentieren und der Maschine mitteilen, mit welchen Ergebnissen sie danebenliegt.

Seit kurzem hat Microsoft seine Suchmaschine auch um die Funktion erweitert Twitter-Nachrichten durchsuchen zu können. Damit ist Microsoft dem Konkurrenten Google einen Schritt voraus, bei dem sich die Twitter-Suche noch in der Erprobungsphase befindet.

Alles in allem macht Bing einen sehr ausgereiften Eindruck, den man aber auch trotz der semantischen Ansätze nicht als bahnbrechenden Fortschritt werten kann. Dennoch konnte Bing einen ordentlichen Start hinlegen und erreichte einen Marktanteil von etwa 6%, was aus den Statistiken des Marktforschungsunternehmens Net Applications hervorging. Die Ziele für Bing fallen derzeit noch sehr zurückhaltend aus: Innerhalb eines Jahres wäre man in Redmond schon damit zufrieden, wenn der Marktanteil überhaupt steigt, so Firmenvertreter.³²

Doch geschlagen geben will sich Microsoft noch lange nicht. „*Google dominiert die Suche und hat in Deutschland sogar 90 Prozent Marktanteil. Aber Google hat die Schlacht nicht gewonnen. Microsoft gibt nicht auf*“³³, sagte der Microsoft-Chef Steve Ballmer.

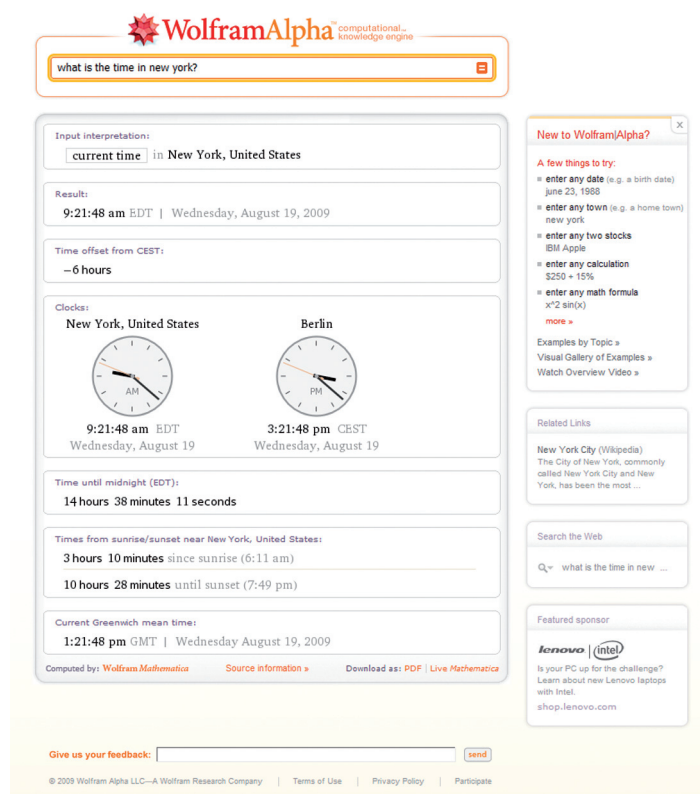
³² Winfuture.de, Bing: Neue Microsoft-Suche erreicht 6% Marktanteil, 03.06.2009, URL: <http://winfuture.de/news,47590.html> Abruf: 19.08.2009

³³ FAZ.net, Web 3.0 – Das Internet der neuen Generation, URL: http://www.faz.net/s/RubE2C6EoBCC2Fo4DD787CDC274993E94C1/Doc~E1628E872A40F41CBB7420E378711B3D6~ATpl~Ecommon~Scontent.html?rss_googlefeed, Abruf: 22.08.2009

3.7 WOLFRAM ALPHA

Seit Mai 2009 ist die Suchmaschine Wolfram Alpha online, die seit 2005 vom Wissenschaftler Stephen Wolfram und einem Team aus rund hundert Mitarbeitern entwickelt wird. Ziel der angeblich neuen „Wunderwaffe“ gegen Google und Co. ist es, Suchanfragen von Menschen zu „verstehen“ und Antworten errechnen zu können. Anders als beispielsweise bei Google ergibt eine Suchanfrage bei Wolfram Alpha keine Liste von Verweisen, deren Text einen oder mehrere Suchbegriffe enthalten. Vielmehr werden bei einer Suchanfrage die passenden Fakten durch spezifische Algorithmen zu Ergebnissen verarbeitet, die dem Benutzer dann in einer übersichtlichen Zusammenfassung nebst Quellenangaben und je nach Thema zusätzlich mit Schaubildern, Diagrammen und Karten dargestellt werden.

Abbildung 7: Suchmaschine Wolfram Alpha und die Suchergebnisse zur Frage „What is the time in New York?“



Wolfram Alpha besteht aus einer gigantischen Datenbank, einer gewaltigen Ansammlung von Algorithmen und einem Übersetzer, der aus sprachlich formulierten Fragen Rechenaufgaben macht, die von den Algorithmen unter Zugriff auf die Datenbasis gelöst werden können. Bei den Daten handelt es sich um „kuratierte“ Daten. Das heißt, dass das meiste Daten-Rohmaterial für die Antworten aus handverlesenen und von Wolfram-Mitarbeitern eigens aufbereiteten Beständen stammt. Und genau hier liegt das Problem. Die Datenbasis von Wolfram Alpha ist noch viel zu schwach um einer Suchmaschine gerecht zu werden, die den Anspruch hat, Google vom Thron zu stoßen. Auch das Versprechen, man könne Fragen stellen, statt nach plumpen Begriffen zu suchen, scheint noch sehr unausgereift. Aktuell bekommt man durchweg brauchbarere Ergebnisse, wenn man Fragestellungen vermeidet. Ein weiterer Kritikpunkt, den sich Wolfram Alpha anschreiben lassen muss, ist der durchgängige Verzicht auf Barrierefreiheit, denn alle Ergebnisse werden als Grafiken ausgegeben, auch reine Texte. Das alles kann sich natürlich noch verbessern im Laufe der Zeit. Das System müsse nur weiterhin ständig gefüttert und ausgebaut werden, was wiederum mit enormen Kosten verbunden ist, da man weiterhin sehr viel Geld in menschliche Arbeitskräfte investieren muss, die neue Daten einpflegen, testen und anpassen. *„Die Frage ist, ob normale Anwender oft genug neue Anläufe starten, um zu erfahren, welche Fragen Wolfram Alpha beantworten kann. Falls nicht, wird es ein wunderbares Werkzeug für Experten bleiben, aber nicht zu dem allgemeinen Allzweckwerkzeug werden“*, bilanziert David Weinberger³⁴. Ein Google-Killer kann man Wolfram Alpha bisher allerdings keinesfalls bezeichnen und das wird sich auch in näherer Zukunft wohl nicht ändern.

³⁴ Spiegel Online, Wolfram Alpha – Hype um den Wissenszwerg, 18.05.2009, URL: <http://www.spiegel.de/netzwelt/web/o,1518,625407,00.html> Abruf: 07.08.2009

3.8 EYEPLORER

„Eine der wohl aufregendsten Suchmaschinen-Entwicklungen der letzten Jahre kommt nicht aus dem Silicon Valley, sondern aus Berlin“³⁵ – so leitet Spiegel Online die Passage über den eyePlover³⁶ ein, in einem Artikel, in dem es eigentlich um Microsofts neue Suchmaschine Bing geht. Entwickelt wurde der eyePlover vom Berliner Startup-Unternehmen Vionto.

Die Vionto GmbH aus Berlin wurde 2008 gegründet und entwickelt laut eigenen Aussagen: „Wissensmaschinen. Maschinen, die wissen. Maschinen, die denken. Maschinen, mit denen man kommunizieren und interagieren kann.“³⁷ Ihr Ansatz orientiert sich dabei an neuesten Erkenntnissen der Kognitionswissenschaften, Computerlinguistik und Neurobiologie.

Die Gründer der Vionto GmbH sind Diplom-Kaufmann und Marketing-Experte Ralf von Grafenstein und der Neurobiologe und Gehirnforscher Dr. Martin C. Hirsch, der Urenkel des Physikers und Nobelpreisträgers Werner Heisenberg.

Auf Basis der Wikipedia-Datenbank stellt der eyePlover thematische Querverbindungen visuell auf einer Scheibe dar. Bei der Entwicklung war es ein wichtiger Ansatz, die assoziativen Fähigkeiten des Menschen auf den eyePlover zu übertragen. Während handelsübliche Suchmaschinen nur verweisen, werden Informationen beim eyePlover visuell aufbereitet und zu Schwerpunkten verknüpft. Im Gegensatz zu ihren Vorgängern ist die Maschine ansatzweise intelligent. Sie findet im Laufe einer Sitzung heraus, worum es dem Benutzer geht, versteht den Zusammenhang von Suche und Inhalt und ist deshalb in der Lage, Empfehlungen auszusprechen.

³⁵ SpiegelOnline, Ballmers Bing bringt den Markt in Bewegung, 18.06.2009, URL: <http://www.spiegel.de/netzwelt/web/o,1518,631220,00.html> Abruf: 23.08.2009

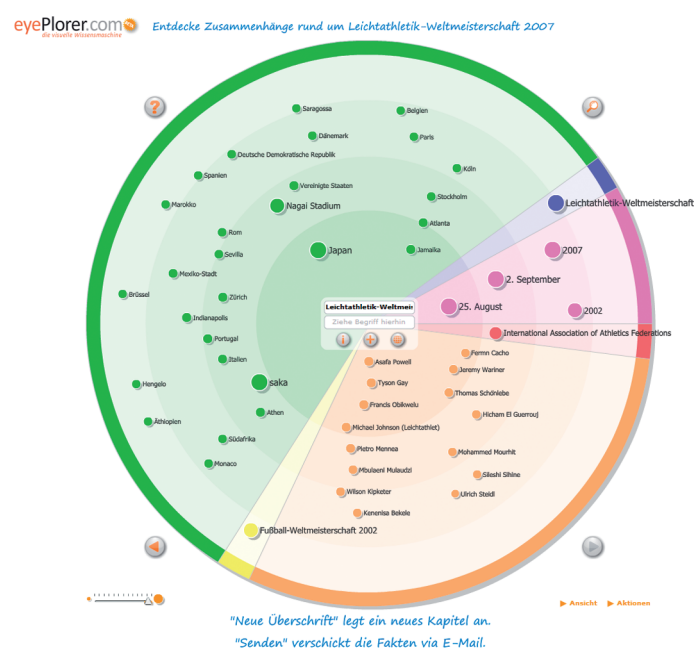
³⁶ URL: <http://www.eyeplover.com>

³⁷ Vionto GmbH, URL: <http://www.vionto.com/2009/01/05/approach/> Abruf: 23.09.2009

Die Erfinder des eyePlorers sprechen daher auch nicht von einer Suchmaschine, sondern vielmehr von einer „Denkmaschine“.

„So etwas für das gesamte Web umzusetzen, statt ‚nur‘ für die Inhalte der Wikipedia, würde dem Traum vom semantischen Web schon sehr nahe kommen.“³⁸

Abbildung 8: visuelle Suchergebnisse beim eyePlorer



Den Lohn einer bis dato erfolgreichen Arbeit konnte Vionto dieses Jahr mit dem Gewinn des Internet-Awards im Bereich „Webportale“ des Verbandes der deutschen Internetwirtschaft „eco“³⁹ bereits einstreichen.

Derzeit versucht Vionto Verlagshäuser und Web-Portale für sich zu gewinnen um einerseits die Wissensbasis des eyePlorers zu erweitern und andererseits die Inhalte in den Portalen selbst besser visualisieren zu können.⁴⁰

³⁸ SpiegelOnline, Ballmers Bing bringt den Markt in Bewegung, 18.06.2009, URL: <http://www.spiegel.de/netzwelt/web/o,1518,631220,00.html> Abruf: 23.08.2009

³⁹ URL: <http://www.eco.de>

⁴⁰ vgl. Gespräch mit Ralf von Grafenstein, dem Gründer der Vionto GmbH zum Thema „eyePlorer“, 27.06.2009

3.9 QIMAYA

Eine weitere Suchmaschine, die in Deutschland entwickelt wird ist Qimaya. Durch die Emulation des menschlichen Gehirns kann Qimaya die Bedeutung respektive den Informationsgehalt von Daten erkennen. Daten können also mittels Künstlicher Intelligenz in einem neuronalen Netz, ähnlich des menschlichen Gehirns, intelligent verknüpft werden. In diesem selbstlernenden System, welches ohne Ontologien und Metadaten arbeitet, entsteht Semantik. Daten kollaborieren selbstständig und automatisch miteinander. Der Fakt, dass Qimaya ohne Metadaten arbeitet, hat einen enormen Vorteil gegenüber anderen semantischen Suchmaschinen, abgesehen vom eyePlorer, der ebenfalls auf Künstliche Intelligenz setzt. So wie jeden Tag neue Inhalte ins Internet gestellt werden, werden sie sich auch ständig ändern, Politiker treten ab, Firmen fusionieren, Institutionen lösen sich auf. Während bei bisherigen Ansätzen semantischer Suchen die Ontologien für diese Fälle neu angelegt werden müssen, „updated“ sich Qimaya quasi von selbst und lernt stetig dazu. Das Qimaya-Team um Dr. Klaus Holthausen sieht aus diesem Grund das Semantic Web, wie es derzeit diskutiert wird, als Idee von gestern an⁴¹. Ganz neu ist die Idee der Verwendung neuronaler Netze jedoch auch nicht, jedoch scheiterten Projekte diesbezüglich immer an der exponentiell steigenden Rechenpower, die dafür aufgebracht werden muss. Ende 2007 gelang diesbezüglich der Durchbruch und man fand eine Lösung für das Problem. Qimaya kann nun also mit dem Bruchteil an Serverkapazität arbeiten.

Momentan werden dabei lediglich 20 Prozent des Normaleinsatzes von Servern eingesetzt. Dadurch wird Qimaya auch erst finanziell umsetzbar, denn es werden so wesentlich weniger Server benötigt, als bisher notwendig gewesen wären.

⁴¹ T3N, Qimaya stellt das Semantic Web auf den Kopf – Das Internet als Gehirn, URL: <http://t3n.yeebase.com/qimaya-stellt-semantic-web-kopf-internet-gehirn-237889/2/>
Abruf: 23.08.2009

Abbildung 9: Qimaya und die Ergebnisliste zum Suchbegriff „Hund“

The screenshot shows the Qimaya search interface. At the top, the Qimaya logo is on the left, and a search bar with the text 'Hund' is in the center. To the right of the search bar is a button labeled 'Qimaya'. Below the search bar is a row of icons representing various semantic categories: a sigma symbol, a person, a building, a dog, a cat, a star, and a magnifying glass. On the left side, there is a vertical list of categories: Hund, Katze, Tiere, Tier, Hunden, Vierbeiner, Hundes, Tierarzt, Haustiere, Fell, Besitzer, Haustier, Futter, Tieres, Leine, Halter, and Rassen. The main area displays a list of search results for 'Hund', each with a title, a URL, and a relevance percentage. The results are as follows:

Category	Result Title	URL	Relevance
Hund	Haustier Magazin - News	www.haustier-magazin.de/hund-katze/news.html	100 %
Katze	Wissenswertes über Hunde Tiere Sat.1 Ratgeber & Magazine	www.sat1.de/ratgeber_magazine/tiere/hunde/content/31828/	98 %
Tiere	ZDF.de - Hitzefrei für Haustiere	www.zdf.de/ZDFde/inhalt/25/0,1872,3945529,00.html	94 %
Tier	Deutscher Tierschutzbund e.V. - FAQ Heimtiere	www.tierschutzbund.de/00524.html	94 %
Hunden	Haustier Magazin - Doppelzimmer mit Hund - Pfötchenhotel für Frauen, Herren und Hund	www.haustier-magazin.de/hund-katze/doppelzimmer-mit-hund-d20070728.html	93 %
Vierbeiner	Planet Wissen - Tiermedizin	www.planet-wissen.de/pw/Artikel.....207CBD49646E223E0440003BA5E092.....html	92 %
Hundes	Haustier Magazin - Tipps für Tierfreunde - Hilfe gegen die Tücken des Alltags	www.haustier-magazin.de/tier-blog/tipper-tierfreunde.html	92 %
Tierarzt	Haustier Magazin - Stress für Haustiere - Weihnachts mit Haustieren	www.haustier-magazin.de/tier-blog/stress-fuer-haustiere.html	91 %
Haustiere	Haustier Magazin - Affenhitze erträglich machen - So bleiben Hund, Katze und Co bei Hitze cool	www.haustier-magazin.de/tier-blog/affenhitze-ertraeglich-machen.html	91 %
Fell	Deutscher Tierschutzbund e.V. - 26.04.07: Urlaubszeit	www.tierschutzbund.de/02583.html	91 %
Besitzer	Haustier Magazin - Tollwut - Hauptüberträger der Tollwut ist der Fuchs	www.haustier-magazin.de/tierlexikon/tollwut.html	90 %
Haustier	NETZEITUNG WISSENSCHAFT NACHRICHTEN: Der beste Freund in Serie	www.netzeitung.de/wissenschaft/1082359.html	90 %
Futter	ZDF.de - Welcher Hund ist der richtige?	www.zdf.de/ZDFde/inhalt/23/0,1872,4087895,00.html	89 %
Tieres	Haustier Magazin - Tier als Weihnachtsgeschenk - (K)ein Tier unter dem Weihnachtsbaum?	www.haustier-magazin.de/tier-blog/tier-als-weihnachtsgeschenk.html	89 %
Leine	gesundheit.de - Zahnygiene nutzt auch Haustieren mit Mundgeruch	www.gesundheit.de/familie/freizeit-zuhause/zahnygiene-haustiere/	89 %

Auf der Abbildung ist zu sehen, dass links von den Ergebnissen eine Reihe von Assoziationsbegriffen zu dem gesuchten Wort angezeigt werden. Dabei können und werden sich die Begriffe ständig ändern und erweitern. Es wird zudem auch die Möglichkeit geben, sich Assoziationsbegriffe zu einem Suchwort aus einem speziellen Jahr anzeigen zu lassen. Desweiteren hat man die Möglichkeit durch die Wahl vorgegebener Kategorien die Suchergebnisse in eine bestimmte Richtung zu lenken. Wer beispielsweise nach „Goethe“ sucht, möchte vielleicht gar kein Buch von Goethe im Internet kaufen, sondern viel lieber ein paar seiner Gedichte lesen. Andersrum kann man natürlich auch eine Kanalisierung von Shoppingangeboten erzwingen. Weiterhin sei noch erwähnt, dass „SEO“⁴² bei Qimaya keinen Erfolg haben wird, denn hier werden die Ergebnisse nicht wie heutzutage üblich nach Prominenz, sondern vielmehr nach Relevanz gefiltert.

⁴² SEO (Search Engine Optimization) ist ein Fachbegriff für Maßnahmen, die dazu dienen die Platzierung einer Internetseite in gängigen Suchdiensten zu verbessern

Derzeit existiert Qimaya lediglich als Betaversion, wird aber laut Aussagen der Entwickler noch 2009 online gehen. Bis dahin werden an einige Neugierige Passwörter herausgegeben, mit denen man sich in die Betaversion einloggen kann, um sie persönlich zu testen.

Ein Problem womit Qimaya anfänglich noch zu kämpfen haben wird, ist der derzeit noch sehr kleine Index, also die verschiedenen Internetseiten, die bisher eingelesen wurden. Das wird sich aber im Laufe der Zeit ändern und Qimaya wird zusehend wachsen, Google hat schließlich auch mal klein angefangen.

Fest steht allerdings, dass Qimaya das Potenzial hat, der Schlüssel zu einem Evolutionsschritt des Internets sein. Dr. Holthausen spricht jetzt schon von einem „*Quantensprung*“⁴³.

⁴³ T3N, Qimaya stellt das Semantic Web auf den Kopf – Das Internet als Gehirn, URL: <http://t3n.yeebase.com/qimaya-stellt-semantic-web-kopf-internet-gehirn-237889/2/>
Abruf: 23.08.2009

4. SEMANTIC WEB AUF DEM WEG IN DEN MASSENMARKT

Momentan spielt das Semantic Web noch eine eher untergeordnete Rolle im Internet. In den vergangenen Jahren haben sich eine Reihe von Startups etabliert, die auf der Grundlage der vom W3C definierten Standards, sowie der Integration unterschiedlicher bedeutungsverarbeitender Softwarekomponenten neue Applikationen und Services auf den Markt gebracht haben. Deren kommerzielle Tragfähigkeit muss sich nun allerdings noch beweisen.

Allerdings haben auf der anderen Seite auch große Player wie Yahoo, Google, Oracle, Sun Microsystems oder die British Telecom „[...] entsprechende Konzepte ohne großes öffentliches Aufsehen bereits internalisiert, wie am Beispiel Google nachzuvollziehen ist.“⁴⁴ Insbesondere in der Medizin und der Medienbranche sind Semantic-Web-Technologien auf dem Vormarsch. Im letzten Fall bekennen sich zum Beispiel Reuters, Dow Jones, die New York Times oder die BBC zum Semantic Web. Auch in Deutschland ist die Medienbranche in Bewegung was die Entwicklung bzw. Implementierung semantischer Technologie betrifft.

Man kann also sagen, dass viele Zeichen darauf hindeuten, dass Semantic-Web-Technologie sukzessive in bereits bestehende Architekturen und Anwendungen integriert werden und somit nach und nach ihren Weg in den Massenmarkt finden werden. Ein weiteres Anzeichen dafür ist das nächste Release des Content Management Systems Drupal, bei dem im Rahmen seiner „Semantic Web Initiative“ eine serienmäßige Integration von RDFa erfolgen wird. Dieser Semantic-Web-Standard für Mikroformate ermöglicht eine maschinenverarbeitbare Annotation von bestimmten Content-Entitäten. Beispielsweise können dann Zahlenwerte automatisch als Preise, als Konfektionsgrößen, als Altersangabe oder als Autokennzeichen ausgewiesen werden.

⁴⁴ Pellegrini 2009, 23 f.

Auch können ganze Dokumenttypen, wie beispielsweise Stellenausschreibungen durch Schemata semantisch angereichert werden und sind somit leichter auffindbar im Netz. Seit einigen Monaten stellt auch Reuters mit seinem Dienst OpenCalais ein automatisches RDF-Verschlagwortungstool für unstrukturierten Text zur Verfügung.

Um den Weg in den Massenmarkt gänzlich zu schaffen muss jedoch beachtet werden, dass das Semantic Web kein Feld für Einzelkämpfer ist, sondern vielmehr ein „[...] *größtechnisches System ist, dessen Relevanz und Qualität ebenso wie andere Netztechnologien von seiner Verbreitung und der gemeinschaftlichen Akzeptanz abhängen.*“⁴⁵

Die Entwicklung entsprechender Geschäftsmodelle schreitet aber auch hier kontinuierlich voran, wobei es wahrscheinlich ist, dass der kommerzielle Erfolg in der ersten Phase vorrangig im B2B-Segment⁴⁶ zu erwarten ist. Hier gibt es bereits jetzt schon Geschäftsmodelle, die auf der Lizenzierung gut strukturierter Metadatenbestände oder auf automatischen „Annotation Services“ für verschiedene Fachthemen, Branchen oder Nutzergruppen aufbauen.

Im B2C-Segment⁴⁷ hingegen nähert sich die Werbewirtschaft über das Schlagwort „kontextualisierte Werbung“ dem Thema Semantik sukzessive an. Dabei steht die Weiterentwicklung der wortbasierten Werbung im Vordergrund. Beim „Concept Advertising“ werden nicht nur der Bedeutungsgehalt eines Wortes, sondern auch sinnvolle Relationen herangezogen, um Werbung sinnvoll zu platzieren. Semantik unterstützt hierbei die Qualitätssicherung von Werbeschaltungen und ermöglicht gleichzeitig die automatisierte Werbewirtschaftung in Nischenthemen.

⁴⁵ Pellegrini 2009, 23 f.

⁴⁶ Business-to-business, Geschäftsbeziehung zwischen mindestens zwei Unternehmen

⁴⁷ Business-to-consumer, Geschäftsbeziehung zwischen Unternehmen und Privatpersonen, bzw. Konsumenten

5. KRITIK UND MACHBARKEIT

Häufig hört man von Kritikern, dass das Semantic Web aus technischer und organisatorischer Sicht viel zu komplex ist, um überhaupt realisiert werden zu können. Dabei stellt das Hauptproblem vieler semantischer Anwendungen den enormen Pflegeaufwand dar. Denn auch wenn das scheinbar intelligente Erkennen von Wörtern und deren Bedeutung künftig die Maschinen übernehmen können, bleibt es trotzdem dem Menschen überlassen, die dafür notwendigen Ontologien festzulegen und entsprechende Metadaten dafür auszuweisen. Da fragt man sich natürlich wer das alles machen soll bei der Fülle von Informationen, die kontinuierlich anwächst.

Noch drastischer drückt es Hirnforscher und eyePlover-Erfinder Dr. Martin C. Hirsch aus: *„Aus meiner Wahrnehmung ist das Semantic Web, wie es beim W3C-Konsortium verankert ist, gescheitert. Zum Einen gibt es 60 Milliarden Seiten im Netz. Kein Mensch kann das in RDF-Tripel, also in ein computerlesbares Format umwandeln. [...]“*⁴⁸

Zudem muss man bedenken, dass Ontologien statisch sind. Das heißt, wenn beispielsweise Firmen fusionieren oder der Deutsche Bundestag neu gewählt wird, können die entsprechend angelegten Ontologien dafür verworfen werden und müssen neu angelegt werden. Des Weiteren wird von einigen Experten davor gewarnt, dass die zunehmende Semantifizierung zu „MetaCrap“ führen könne, der erst recht keine Verbesserung von Suche und Datenintegration erlaubt und die erhofften Effizienzgewinne und Usability-Verbesserungen letzten Endes wieder konterkariert. Weiterhin bestände die Gefahr, dass durch irreführende Annotationen die Spam-Problematik in eine neue Ebene gehievt werden könnte. Ebenso unklar ist inwieweit sich in einem Semantic Web die Anforderungen an Datenschutz, Urheberrecht und Wettbewerbsregulierung ändern werden. Experten attestieren an dieser Stelle allerdings schon Handlungsbedarf auf gesetzgebender Ebene.

⁴⁸ Debug, 134 Juli / August, 22 ff.

Neben der bis dato fehlenden Technologie für die Umsetzung eines Semantic Web, die nun vorhanden ist, war auch angesichts der exponentiell steigenden Datenmenge, die fehlende Serverkapazität bisher ein Problem gewesen und ließ semantische Anwendungen dadurch finanziell nicht tragbar machen. Allerdings arbeitet hier die exponentiell ansteigende technische Entwicklung dagegen, welche das Problem künftig beseitigen wird.

Die genannte Kritik darf keinesfalls auf die leichte Schulter genommen werden, denn sie bringt nicht nur die berechtigte Skepsis, sondern auch die genannten Problemfelder zum Ausdruck. Außerdem relativiert sie die teilweise unreflektierte Technologie-Euphorie und den Machbarkeitsfetisch der Technik-Geeks.

Bei aller berechtigter Kritik, muss an dieser Stelle jedoch gesagt werden, dass es auch semantische Technologien wie beispielsweise Qimaya oder den eyePlorer gibt, die ohne Annotationen und Ontologien auskommen werden, also tatsächlich auf Künstlicher Intelligenz beruhen und damit bereits einem Großteil der genannten Kritik entgegenwirken.

6. FAZIT

Die Semantifizierung des Internets ist bereits im Gange, wird aber noch sehr lange andauern, wenn man bedenkt dass die Grundlage der meisten semantischen Anwendungen eine Integrierung von Ontologien und Metadaten in möglichst allen relevanten Internetseiten nötig macht. Für neu erstellte Internetseiten wird es spezielle Editoren geben, in die man semantische Tags eingibt. Alle bestehenden Internetseiten sollen mittels einer speziellen Software automatisch in semantische umgewandelt werden. Eine andere Möglichkeit ist die Nutzerpartizipation. Das THESEUS-Programm will die Internetgemeinde auffordern, sich am „Rückwärts-Taggen“ zu beteiligen. Laut Projektbeschreibung sollen Nutzer mit Hilfe eines semantischen Werkzeugkastens *„selbst Inhalte, Regeln und Ordnungen erstellen und bearbeiten sowie multimediale Inhalte intelligent aufbereiten, sammeln und verknüpfen können“*⁴⁹.

Wie es besser geht, machen die derzeit wahrscheinlich innovativsten semantischen Webdienste Qimaya und eyePlorer vor, die auf Grundlage Künstlicher Intelligenz funktionieren und Ontologien und Metadaten damit überflüssig machen.

Mittlerweile gibt es schon zahlreiche semantische Webanwendungen, die sich jedoch alle noch etablieren müssen. Besonders bei den „Google-Killern“, wie manche semantische Suchdienste sich gerne selbst bezeichnen oder von der Presse bezeichnet werden, wird sich zeigen, ob die breite Masse der Nutzer bei der Suche nach Informationen zukünftig wirklich mehr auf relevante statt prominente Suchergebnisse à la Google setzen wird. Wobei man hier aber erwähnen muss, dass auch Google an semantischen Funktionen seiner Suche arbeitet und auch schon teilweise integriert hat. Man kann aber davon ausgehen, dass sich hinter den Kulissen des Weltkonzerns einiges mehr tut, als man bisher zugeben will. Man kann also gespannt sein, ob Google in den nächsten Jahren kontern wird.

⁴⁹ vgl. manager-magazin.de, Semantic Web - Das klügere Netz, URL: <http://www.manager-magazin.de/it/artikel/o,2828,563728-2,00.html> Abruf: 26.08.2009

Häufig wird das Semantic Web in Zusammenhang mit dem Web 2.0 erwähnt. Während beim Web 2.0 allerdings Massenphänomene wie „folksonomy“⁵⁰, „user generated content“⁵¹, oder „crowdsourcing“⁵² im Vordergrund stehen, verfolgt das Semantic Web das Ziel, die Qualität vorhandener Informationen auf der semantischen Ebene zu verbessern. Eine Zusammenführung der Technologie des Semantischen Web mit sozialen Ansätzen des Web 2.0 wird oft auch als „Social Semantic Web“⁵³ oder als „Web 3.0“ bezeichnet. Dass wir auf dem Weg zum Web 3.0 bereits sind, kann keiner mehr leugnen, wann wir es erreichen kann man nicht genau sagen, da der Übergang hier fließend erfolgt.

Aber wird es nun irgendwann möglich sein bei einer Suchmaschine mit einer konkreten Fragestellung nach dem nächsten Blumenladen in der Nähe zu suchen, der rote Rosen auf Lager hat? Denkbar ist das auf jeden Fall. Konkrete Fragestellungen zu verstehen funktioniert bei einigen semantischen Anwendungen wie beispielsweise Hakia heute schon recht gut, wenn auch nur im Ansatz. Auch die Komplexität der hier als Beispiel aufgeführten Frage übersteigt das „Wissen“ der Maschinen momentan noch. Die zur Beantwortung notwendige Abfrage, wo der Suchende sich momentan aufhält, funktioniert allerdings mittels IP-Lokalisation schon jetzt. Ob diese Funktion irgendwann der breiten Masse zugänglich gemacht wird, ist allerdings begründet durch Datenschutz und Co. sehr zweifelhaft. Allerdings wäre hier vorstellbar, dass man seine Postleitzahl mit der Suchanfrage angibt oder seine Adressdaten vorher in einem Profil speichert.

⁵⁰ Begriff aus dem Web 2.0, steht für eine Sammlung von verschlagworteten Begriffen

⁵¹ steht für Inhalte, die nicht vom Anbieter eines Webangebots, sondern von dessen Nutzern erstellt werden

⁵² bezeichnet die Auslagerung auf die Intelligenz und die Arbeitskraft einer Masse von Freizeitarbeitern im Internet

⁵³ vgl. Blumauer A./Pellegrini 2009, 88

Besonderes Augenmerk gilt auch den verschiedenen Eingabemethoden. Beim semantischen Web sollen Informationen sowohl per Mobilgerät und Spracheingabe abgerufen werden können, als auch durch visuellen Input. Wer ein Filmplakat mit dem Handy abfotografiert und zusätzlich „Wo läuft dieser Film?“ ins Mikro spricht, soll in ferner Zukunft eine Liste mit nahegelegenen Kinos zurückbekommen. Dabei wird das so genannte „Geo-Tagging“ verwendet, damit das Handy dann auch weiß, wo das nächste Kino zu finden ist.

Welche Auswirkungen das Semantic Web nun letztendlich haben wird, wird sich jedoch erst viel später zeigen, die Möglichkeiten allerdings sind atemberaubend. Einige Verfechter des Semantic Web behaupten sogar, dass es zu einer Evolution des menschlichen Wissens selbst führen wird, da es dem Menschen dadurch zum ersten Mal möglich sein wird, die Unmengen der auf dieser Welt vorhandenen Daten auf relevante und produktive Art zu filtern und zu kombinieren.

Fakt ist, dass wir die bereitstehenden Informationen im Internet ohne das Semantic Web in ein paar Jahren nicht mehr effektiv nutzen können, da wir einem regelrechten „Informations-Overload“ ausgesetzt sein werden, in dem niemand mehr durchsieht. Um der stetig wachsenden Fülle an Informationen Herr zu werden, brauchen wir ein semantisches Internet.

7. LITERATURVERZEICHNIS

7.1 BÜCHER

Blumauer A./Pellegrini, T (Hrsg.): Social Semantic Web. Web 2.0 - was nun? Heidelberg 2009

Hitzler, P./Krötzsch, M./Rudolph, S. et al.: Semantic Web, Berlin 2008

7.2 ZEITSCHRIFTEN

Debug, 134 Juli / August, 22 ff.

Pellegrini, Tassilo: Grundlagen des Semantic Web – Fortschritt mit (r)evolutionärem Potenzial In: T3N-Magazin, Nr. 14, 23 f.

7.3 ZEITUNGEN

die Tageszeitung, Nr. 8862 / 17.04.2009, 13

Frankfurter Allgemeine Zeitung, 21.02.2009, o.S.

7.4 ONLINEQUELLEN

Bing, URL: <http://www.bing.com>, Abruf: 19.08.2009

Bundesministerium für Wirtschaft und Technologie, THESEUS-Basistechnologien, <http://theseus-programm.de/basistechnologien/default.aspx> Abruf: 12.08.2009

Bundesministerium für Wirtschaft und Technologie, Was ist THESEUS? <http://theseus-programm.de/was-ist-theseus/default.aspx> Abruf: 12.08.2009

Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., Fast jeder fünfte Mensch ist online, URL: http://www.bitkom.org/de/presse/49919_46069.aspx, Abruf: 26.08.2009

Bush, Vannevar 1945, As We May Think, URL: <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml>, Abruf: 26.08.2009

Chip Online, Bing: 15 geniale Tipps für die Microsoft-Suche, URL: http://www.chip.de/bildergalerie/Bing-15-geniale-Tipps-fuer-die-Microsoft-Suche-Galerie_36788597.html, Abruf: 19.08.2009

Computerwoche, Microsoft kauft Powerset, <http://www.computerwoche.de/netzwerke/web/1867840/>, Abruf: 19.08.2009

Evri.com, Angela Merkel, <http://www.evri.com/person/angela-merkel-ox2627d>, Abruf: 16.08.2009

FAZ.net, Web 3.0 – Das Internet der neuen Generation, URL: http://www.faz.net/s/RubE2C6EoBCC2Fo4DD787CDC274993E94C1/Doc~E1628E872A40F41CBB7420E378711B3D6~ATpl~Ecommon~Scontent.html?rss_googlefeed, Abruf: 19.08.2009

Focus Online, Bing: Microsoft startet neue Suchmaschine, 28.05.2009,
URL: http://www.focus.de/digital/internet/bing-microsoft-startet-neue-suchmaschine_aid_403530.html, Abruf: 19.08.2009

Hakia-Lab, Query Detection and Extraction – QDEX - System <http://labs.hakia.com/hakia-lab-qdex.html> Abruf: 19.08.2009

Heise Online, Bundesregierung will am semantischen Suchinstrument Theseus festhalten, <http://www.heise.de/newsticker/Bundesregierung-will-am-semantischen-Suchinstrument-Theseus-festhalten--meldung/87220> Abruf: 12.08.2009

Heise.de, Microsoft übernimmt Suchspezialisten Powerset, <http://www.heise.de/newsticker/Microsoft-uebernimmt-Suchspezialisten-Powerset--meldung/110327>, Abruf: 19.08.2009

Heise Online, Organizer für das semantische Web, <http://www.heise.de/tr/Organizer-fuer-das-semantische-Web--artikel/118218> Abruf: 12.08.2009

Heise Online, Semantische Suchmaschine gestartet, <http://www.heise.de/newsticker/Semantische-Suchmaschine-gestartet--meldung/107759>, Abruf: 19.08.2009

Heise Online, Theseus soll dem Arzt helfen, <http://www.heise.de/newsticker/Theseus-soll-dem-Arzt-helfen--meldung/118793> Abruf: 12.08.2009

Heise Online, Wolfram Alpha ist online, <http://www.heise.de/newsticker/Wolfram-Alpha-ist-online-Update--meldung/137963> Abruf: 07.08.2009

manager-magazin.de, Neue Suchmaschine – Griff nach Googles heiligem Gral, <http://www.manager-magazin.de/it/artikel/0,2828,471830-3,00.html>, Abruf: 19.08.2009

manager-magazin.de, Semantic Web - Das klügere Netz, URL: <http://www.manager-magazin.de/it/artikel/o,2828,563728-2,00.html> Abruf: 26.08.2009

Pseudopost.org, Web 4.0, NLP und hakia vs. Google, <http://pseudopost.org/archives/2007/03/web-40-nlp-und-hakia-vs-google/> Abruf: 19.08.2009

Qimaya, Qimaya - Semantic Web - Immer her mit der Kritik, URL: <http://qimaya.wordpress.com/2008/10/02/immer-her-mit-der-kritik/> Abruf: 23.08.2009

SEO-News.at, Qimaya, die intelligente Suchmaschine, URL: <http://www.seo-news.at/qimaya-die-intelligente-suchmaschine>, Abruf: 23.08.2009

Spiegel Online, Ballmers Bing bringt den Markt in Bewegung, 18.06.2009, URL: <http://www.spiegel.de/netzwelt/web/o,1518,631220,00.html>, Abruf: 23.08.2009

Spiegel Online, Internet-Krimi: David gegen Google, <http://www.spiegel.de/wirtschaft/o,1518,471577,00.html>, Abruf: 19.08.2009
Golem.de, Powerset startet semantische Suchmaschine, <http://www.golem.de/o805/59655.html>, Abruf: 19.08.2009

Spiegel Online, Microsofts Suchmaschine Bing – Praktisch, treffsicher, Google-gut, 03.06.2009, URL: <http://www.spiegel.de/netzwelt/web/o,1518,628146,00.html>, Abruf: 19.08.2009

Spiegel Online, Physiker Wolfram – Software-Genie Wolfram verspricht den Google-Killer, <http://www.spiegel.de/netzwelt/tech/o,1518,612268,00.html> Abruf: 09.08.2009

Spiegel Online, Suchmaschine Wolfram Alpha – Was der Google-Gegner weiß – Und was nicht, <http://www.spiegel.de/netzwelt/web/o,1518,623122,00.html> Abruf: 09.08.2009

Spiegel Online, Wolfram Alpha – Hype um den Wissenszwerg, <http://www.spiegel.de/netzwelt/web/o,1518,625407,00.html> Abruf: 07.08.2009

T3N-Magazin, Evri – semantische Verknüpfung verschiedener Inhalte, <http://t3n.yeebase.com/evri-semantische-verknupfung-verschiedener-webinhalte-237044/> Abruf: 16.08.2009

T3N, Qimaya stellt das Semantic Web auf den Kopf – Das Internet als Gehirn, URL: <http://t3n.yeebase.com/qimaya-stellt-semantic-web-kopf-internet-gehirn-237889/2/> Abruf: 23.08.2009

@Web Suchmaschinen Weblog, Suchmaschine Hakia – Ergebnisse vertrauenswürdig und glaubhaft, <http://www.at-web.de/blog/20081007/suchmaschine-hakia-ergebnisse-vertrauenswurdig-und-glaubhaft.htm>, Abruf: 19.08.2009

Winfuture.de, Bing: Microsoft stellt die ‚Entscheidungsmaschine‘ vor, 28.05.2009, URL: <http://winfuture.de/news,47503.html>, Abruf: 19.08.2009

Winfuture.de, Bing: Neue Microsoft-Suche erreicht 6% Marktanteil, 03.06.2009, URL: <http://winfuture.de/news,47590.html>, Abruf: 19.08.2009

7.5 PERSÖNLICHE GESPRÄCHE

Gespräch mit Ralf von Grafenstein, einem der Gründer der Vionto GmbH zum Thema „eyePlorer“, 27.06.2009

SELBSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Alle Teile, die wörtlich oder sinngemäß einer Veröffentlichung entstammen, sind als solche kenntlich gemacht. Die Arbeit wurde noch nicht veröffentlicht oder einer anderen Prüfungsbehörde vorgelegt.

Berlin, den 27.08.2009

Sebastian Schrama

